Bounds for Integral j-Invariants and Cartan Structures on Elliptic Curves

Yuri Bilu, Pierre Parent (Université de Bordeaux I)

July 15, 2008

Abstract

We bound the j-invariant of integral points on a modular curve in terms of the congruence group defining the curve. We apply this to prove that the modular curve $X_{\mathrm{split}}(p^3)$ has no non-trivial rational point if p is a sufficiently large prime number. Assuming the GRH, one can replace p^3 by p^2.

AMS 2000 Mathematics Subject Classification 11G18 (primary), 11G05, 11G16 (secondary).

1 Introduction

Let $N \geq 2$ be an integer and $X(N)$ the principal modular curve of level N. Further, let G a subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ and let X_G be the corresponding modular curve. This curve is defined over $\mathbb{Q}(\zeta_N)^{\det(G)}$, so in particular it is defined over $\mathbb{Q}$ if $\det(G) = (\mathbb{Z}/N\mathbb{Z})^\times$. (Through all this paper, we say that an algebraic curve is *defined* over a field if it has a geometrically integral model over this field.) As usual, we denote by Y_G the finite part of X_G (that is, X_G deprived of the cusps). If X_G is defined over a number field K, the curve X_G has a natural (modular) model over $\mathcal{O} = \mathcal{O}_K$ that we still denote by X_G. The cusps define a closed subscheme of X_G over $\mathcal{O}$, and we define the relative curve Y_G over $\mathcal{O}$ as X_G deprived of the cusps. If S is a finite set of places of K containing the infinite places, then the set of S-integral points $Y_G(\mathcal{O}_S)$ consists of those $P \in Y_G(K)$ for which $j(P) \in \mathcal{O}_S$, where j is, as usual, the modular invariant and $\mathcal{O}_S = \mathcal{O}_{K,S}$ is the ring of S-integers.

In its simplest form, the first principal result of this article gives an explicit upper bound for $j(P) \in \mathbb{Z}$ under certain Galois condition on the cusps. More precisely, we prove the following.

Theorem 1.1 *Assume that X_G is defined over $\mathbb{Q}$, and assume that the absolute Galois group $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ acts non-transitively on the cusps of X_G. Then for any $P \in Y_G(\mathbb{Z})$ we have*

$$\log |j(P)| \leq 30|G|N^2 \log N. \tag{1}$$

This result was announced in [1]. Because of an inaccuracy in the proof given in [1], the $\log N$ factor is missing therein (which, however, does not affect the arithmetical applications of this theorem).

Actually, we obtain two versions of Theorem 1.1. One (Theorem 1.2 below) is quite general, applies to any number field and a ring of S-integers in it, but the bound is slightly weaker. The other (see Section 7) is less general than Theorem 1.1, and applies only to certain particular groups G (the normalizers of split tori), but the bound is sharper.

To state Theorem 1.2, we need to introduce some notation. We denote by $\mathrm{h}(\cdot)$ the usual absolute logarithmic height (see Subsection 1.1). For $P \in X_G(\bar{\mathbb{Q}})$ we shall write $\mathrm{h}(P) = \mathrm{h}\big(j(P)\big)$. For a number field K we denote by $\mathcal{C} = \mathcal{C}(G)$ the set of cusps of X_G, and by $\mathcal{C}(G, K)$ the set of $\mathrm{Gal}(\bar{K}/K)$-orbits of $\mathcal{C}$.

Theorem 1.2 *Let K be a number field and S a finite set of places of K (including all the infinite places). Let G be a subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ such that X_G is defined over K. Assume that $|\mathcal{C}(G,K)| > |S|$ (the "Runge condition"). Then for any $P \in Y_G(\mathcal{O}_S)$ we have*

$$\mathrm{h}(P) \le s^{s/2+1}\left(|G|N^2\right)^s N(\mathcal{R}+30), \tag{2}$$

where $s = |S|$ and

$$\mathcal{R} = \mathcal{R}(N,S) = \sum_{\substack{p \mid N \\ v \mid p \ \text{for some} \ v \in S}} \frac{\log p}{p-1}, \tag{3}$$

the sum being over all the prime divisors of N below the (finite) places from S (in particular $\mathcal{R} = 0$ if S consists only of infinite places).

While this theorem applies in the set-up of Theorem 1.1, it implies a slightly weaker result, with $30|G|N^3$ on the right. Mention also that the constant 30 is not best possible for the method and can be easily reduced, at least for large N.

These theorems are proved in Sections 5 and 6 by a variation of the method of Runge, after some preparation in Sections 2, 3 and 4. In Section 7 we obtain an especially sharp version of Theorem 1.1 for the case of split tori. For a general discussion of Runge's method see [2, 10].

In Section 8 we apply these results to the arithmetic of elliptic curves. We are motivated by a question of Serre, who proved [19] that for any elliptic curve E without complex multiplication (CM in the sequel), there exists a constant $p_0(E)$ such that for every prime $p > p_0(E)$ the natural Galois representation $\rho_{E,p} : \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(E[p]) \cong \mathrm{GL}_2(\mathbb{F}_p)$ is surjective. Masser and Wüstholz [13] gave an effective version of Serre's result; see also the more recent work of Cojocaru and Hall [3, 4].

Serre asks whether p_0 can be made independent of E:

> *does there exist an absolute constant p_0 such that for any non-CM elliptic curve E over $\mathbb{Q}$ and any prime $p > p_0$ the Galois representation $\rho_{E,p}$ is surjective?*

The general guess is that $p_0 = 37$ would probably do.

We obtain several results on Serre's question. One knows that, for a positive answer, it is sufficient to bound the primes p such that a non-CM curve may have a Galois structure included in the normalizer of a (split or nonsplit) Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. Equivalently, one would like to prove that, for large p, the only rational points of the modular curves $X_{\mathrm{split}}(p)$ and $X_{\mathrm{nonsplit}}(p)$ are the cusps and CM points, in which case we will say that the rational points are *trivial* (for the precise definition of these curves see Section 7). In [16, 18] it was proved, by very different techniques, that $X_{\mathrm{split}}(p)(\mathbb{Q})$ is trivial for a (large) positive density of primes; but the methods of loc. cit. have failed to prevent a complementary set of primes from escaping them. Here we consider Cartan structures modulo some higher power of a prime, and we prove the following.

Theorem 1.3 *For large enough prime p, every point in $X_{\mathrm{split}}(p^3)(\mathbb{Q})$ is either a CM-point or a cusp. Assuming the Generalized Riemann Hypothesis for the zeta functions of number fields (GRH in the sequel), the same holds true for $X_{\mathrm{split}}(p^2)(\mathbb{Q})$.*

Equivalently, for large enough p and for any non-CM elliptic curve E defined over $\mathbb{Q}$, the image of the Galois representation ρ_{E,p^3} is not contained in the normalizer of a split Cartan subgroup of $\mathrm{GL}_2(\mathbb{Z}/p^3\mathbb{Z})$ (and p^3 can be replaced by p^2 assuming GRH).

The second part, where one assumes GRH, was sketched in [1], with level p^5. Here we manage to reduce to level p^2 thanks to the refined bound from Section 7, and we prove the first, unconditional assertion of Theorem 1.3 by applying the isogeny estimate of Masser and Wüstholz [11], made explicit by Pellarin [17].

Call a prime number p *deficient* for an elliptic curve E if $\rho_{E,p}$ is not surjective; we call p a *(non-)split Cartan deficient prime* if the image of $\rho_{E,p}$ is contained in the normalizer of a (non) split Cartan subgroup. Following a suggestion of L. Merel and J. Oesterlé, we prove that the split Cartan deficiencies are bounded, with at most 2 exceptions.

Theorem 1.4 *There exists an absolute effective constant p_0 such that for any non-CM elliptic curve $E/\mathbb{Q}$, all but 2 split Cartan deficient primes do not exceed p_0.*

Acknowledgments We thank Daniel Bertrand, Henri Cohen, Loïc Merel, Joseph Oesterlé, Vinayak Vatsal and Yuri Zarhin for stimulating discussions and useful suggestions. We specially acknowledge that the idea of Theorem 1.4 came out in a conversation with Merel and Oesterlé.

1.1 Notation, conventions

Everywhere in this article log and arg stand for the principal branches of the complex logarithm and argument functions; that is, for any $z \in \mathbb{C}^\times$ we have $-\pi < \operatorname{Im} \log z = \arg z \le \pi$. We shall systematically use, often without special reference, the estimates of the kind

$$\begin{aligned} |\log(1+z)| \le \frac{|\log(1-r)|}{r}|z|, \qquad |e^z - 1| \le \frac{e^r-1}{r}|z|, \\ \left|(1+z)^A - 1 - Az\right| \le \frac{\left|1+\varepsilon r)^A - 1 - \varepsilon Ar\right|}{r^2}|z|^2 \quad (\varepsilon = \operatorname{sign} A), \end{aligned} \tag{4}$$

etc., for $|z| \le r < 1$. They can be easily deduced from the Schwarz lemma.

Let $\mathcal{H}$ denote the upper half-plane of the complex plane: $\mathcal{H} = \{\tau \in \mathbb{C} : \operatorname{Im} \tau > 0\}$. For $\tau \in \mathcal{H}$ we put $q_\tau = e^{2\pi i \tau}$. We put $\bar{\mathcal{H}} = \mathcal{H} \cup \mathbb{Q} \cup \{i\infty\}$. If Γ is a pull-back of $G \cap \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ to $\mathrm{SL}_2(\mathbb{Z})$, then the set $X_G(\mathbb{C})$ of complex points is analytically isomorphic to the quotient $X_\Gamma = \bar{\mathcal{H}}/\Gamma$, supplied with the properly defined topology and analytic structure [9, 21].

We denote by D the standard fundamental domain of $\mathrm{SL}_2(\mathbb{Z})$ (the hyperbolic triangle with vertices $e^{\pi i/3}$, $e^{2\pi i/3}$ and $i\infty$, together with the geodesic segments $[i, e^{2\pi i/3}]$ and $[e^{2\pi i/3}, i\infty]$). Notice that for $\tau \in D$ we have $|q_\tau| \le e^{-\pi\sqrt{3}} < 0.005$, which will be systematically used without special reference.

For $\mathbf{a} = (a_1, a_2) \in \mathbb{Q}^2$ we put $\ell_{\mathbf{a}} = B_2\bigl(a_1 - \lfloor a_1 \rfloor\bigr)$ where $B_2(T) = T^2 - T + 1/6$ is the second Bernoulli polynomial. The quantity $\ell_{\mathbf{a}}$ is $\mathbb{Z}^2$-periodic in $\mathbf{a}$ and is thereby well-defined for $\mathbf{a} \in (\mathbb{Q}/\mathbb{Z})^2$ as well: for such $\mathbf{a}$ we have $\ell_{\mathbf{a}} = B_2(\widetilde{a}_1)$, where $\widetilde{a}_1$ is the lifting of the first coordinate of $\mathbf{a}$ to the interval $[0, 1)$. Obviously, $|\ell_{\mathbf{a}}| \le 1/12$; this will also be often used without special reference.

We fix, once and for all, an algebraic closure $\bar{\mathbb{Q}}$ of $\mathbb{Q}$, which is assumed to be a subfield of $\mathbb{C}$. In particular, for every $a \in \mathbb{Q}$ we have the well defined root of unity $e(a) = e^{2\pi i a} \in \bar{\mathbb{Q}}$. Every number field used in this article is presumed to be contained in the fixed $\bar{\mathbb{Q}}$. If K is such a number field and v is a valuation on K, then we tacitly assume than v is somehow extended to $\bar{\mathbb{Q}} = \bar{K}$; equivalently, we fix an algebraic closure $\bar{K}_v$ and an embedding $\bar{\mathbb{Q}} \hookrightarrow \bar{K}_v$. In particular, the roots of unity $e(a)$ are well-defined elements of $\bar{K}_v$.

For a number field K we denote by M_K the set of all valuations (or places) of K normalized to extend the usual infinite and p-adic valuations of $\mathbb{Q}$: $|2|_v = 2$ if $v \in M_K$ is infinite, and $|p|_v = p^{-1}$ if v extends the p-adic valuation of $\mathbb{Q}$. In the finite case we sometimes use the additive notation $v(\cdot)$, normalized to have $v(p) = 1$. We denote by M_K^∞ and M_K^0 the subsets of M_K consisting of the infinite (archimedean) and the finite (non-archimedean) valuations, respectively.

Recall the definition of the absolute logarithmic height $\mathrm{h}(\cdot)$. For $\alpha \in \bar{\mathbb{Q}}$ we pick a number field K containing α and put $\mathrm{h}(\alpha) = [K:\mathbb{Q}]^{-1} \sum_{v \in M_K} [K_v : \mathbb{Q}_v] \log^+ |\alpha|_v$, where the valuations on K are normalized to extend standard infinite and p-adic valuations on $\mathbb{Q}$ and $\log^+ x = \log \max\{x, 1\}$. The value of $\mathrm{h}(\alpha)$ is known to be independent on the particular choice of K. As usual, we extend the definition of the height to $\mathbb{P}^1(\bar{\mathbb{Q}}) = \bar{\mathbb{Q}} \cup \{\infty\}$ by setting $\mathrm{h}(\infty) = 0$. If α is a rational integer or an imaginary quadratic integer then $\mathrm{h}(\alpha) = \log |\alpha|$.

2 Estimates for Modular Functions at Infinity

The results of this section must be known, but we did not find them in the available literature, so we state and prove them here. Most of the results of this section are stronger than what we actually need, but we prefer to state them in this sharp form for the sake of further applications.

2.1 Estimating the j-Function

Recall that the modular j-invariant $j : \mathcal{H} \to \mathbb{C}$ is defined by $j(\tau) = (12c_2(\tau))^3/\Delta(\tau)$, where

$$c_2(\tau) = \frac{(2\pi i)^4}{12}\left(1 + 240\sum_{n=1}^{\infty}\frac{n^3 q_\tau^n}{1-q^n}\right)$$

(see, for instance, [8, Section 4.2]) and $\Delta(\tau) = (2\pi i)^{12} q \prod_{n=1}^{\infty}(1-q^n)^{24}$ (here $q = q_\tau = e^{2\pi i\tau}$). Also, j has the familiar Fourier expansion $j(\tau) = q^{-1} + 744 + 196884q + \ldots$.

Proposition 2.1 *For $\tau \in \mathcal{H}$ such that $|q_\tau| \le 0.005$ (and, in particular, for every $\tau \in D$) we have*

$$\left|j(\tau) - q_\tau^{-1} - 744\right| \le 330000|q_\tau|. \tag{5}$$

(Recall that D is the standard fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$.)

Proof We write $q = q_\tau$. Using the estimate $n^3 \le 3^n$ for $n \ge 3$, we find that for $|q| < 1/3$

$$\begin{aligned}\left|\frac{12}{(2\pi i)^4}c_2(\tau) - 1 - 240q\right| &\le 240\left(\frac{|q|^2}{1-|q|} + \sum_{n=2}^{\infty}\frac{n^3|q|^n}{1-|q|^n}\right)\\ &\le \frac{240}{1-|q|}\left(|q|^2 + 8|q|^2 + \sum_{n=3}^{\infty}|3q|^n\right)\\ &= \frac{2160}{(1-|q|)(1-3|q|)}|q|^2,\end{aligned}$$

and for $|q| \le 0.005$ we obtain

$$\left|\frac{12}{(2\pi i)^4}c_2(\tau) - 1 - 240q\right| \le 2204|q|^2. \tag{6}$$

Further, using (4), we obtain, for $|q| \le 0.005$,

$$\left|\log\frac{(2\pi i)^{12}q(1-q)^{24}}{\Delta(\tau)}\right| = 24\left|\sum_{n=2}^{\infty}\log(1-q^n)\right| \le 24.1\sum_{n=2}^{\infty}|q|^n \le 24.3|q|^2.$$

Hence

$$\begin{aligned}\left|\frac{(2\pi i)^{12}q}{\Delta(\tau)} - 1 - 24q\right| &\le \left|(1-q)^{-24}\right|\left|\frac{(2\pi i)^{12}q(1-q)^{24}}{\Delta(\tau)} - 1\right| + \left|(1-q)^{-24} - 1 - 24q\right|\\ &\le 1.13\left|\log\frac{(2\pi i)^{12}q(1-q)^{24}}{\Delta(\tau)}\right| + 314|q|^2 \le 342|q|^2.\end{aligned}$$

Combining this with (6), we obtain (5) after a tiresome, but straightforward calculation. ☐

Corollary 2.2 *For any $\tau \in D$ we have either $|j(\tau)| \le 2500$ or $|q_\tau| < 0.001$.* ☐

2.2 Estimating Siegel's Functions

For a rational number a we define $q^a = e^{2\pi i a\tau}$. Let $\mathbf{a} = (a_1, a_2) \in \mathbb{Q}^2$ be such that $\mathbf{a} \notin \mathbb{Z}^2$, and let $g_{\mathbf{a}} : \mathcal{H} \to \mathbb{C}$ be the corresponding *Siegel function* [7, Section 2.1]. Then, putting $z = a_1\tau + a_2$ and $q_z = q_\tau^{a_1} e(a_2)$, where $e(a) = e^{2\pi i a}$, we have the following infinite product presentation for $g_{\mathbf{a}}$ [7, page 29] (where $B_2(T)$ is the second Bernoulli polynomial):

$$g_{\mathbf{a}}(\tau) = -q_\tau^{B_2(a_1)/2} e\left(\frac{a_2(a_1-1)}{2}\right)(1-q_z)\prod_{n=1}^{\infty}(1-q_\tau^n q_z)(1-q_\tau^n/q_z). \tag{7}$$

We also have [7, pages 29–30] the relations

$$g_{\mathbf{a}} \circ \gamma = g_{\mathbf{a}\gamma} \cdot (\text{a root of unity}) \quad \text{for} \quad \gamma \in \Gamma(1), \tag{8}$$

$$g_{\mathbf{a}} = g_{\mathbf{a}'} \cdot (\text{a root of unity}) \quad \text{when} \quad \mathbf{a} \equiv \mathbf{a}' \mod \mathbb{Z}^2. \tag{9}$$

Remark that the roots of unity in (8) and (9) are of order dividing $12N$; this will be used later.

The order of vanishing of $g_{\mathbf{a}}$ at $i\infty$ (that is, the only rational number ℓ such that the limit $\lim\limits_{\tau\to i\infty} q_\tau^{-\ell} g_{\mathbf{a}}(\tau)$ exists and is non-zero) is equal to the number $\ell_{\mathbf{a}}$, defined in Subsection 1.1, see [7, page 31].

Proposition 2.3 *Assume that $a_1 \notin \mathbb{Z}$ and let N be a denominator of a_1 (that is, a positive integer satisfying $Na_1 \in \mathbb{Z}$). Then for $|q_\tau| \le 10^{-N}$ we have*

$$\Bigl|\log|g_{\mathbf{a}}(\tau)| - \ell_{\mathbf{a}}\log|q_\tau|\Bigr| \le 3|q_\tau|^{1/N}. \tag{10}$$

Further, assume that $a_1 \in \mathbb{Z}$. Then for $|q_\tau| \le 0.1$ we have

$$\Bigl|\log|g_{\mathbf{a}}(\tau)| - \ell_{\mathbf{a}}\log|q_\tau| - \log|1-e(a_2)|\Bigr| \le 3|q_\tau|. \tag{11}$$

Proof Due to (8), we may assume that $0 \le a_1 < 1$ and distinguish between the cases $0 < a_1 < 1$ and $a_1 = 0$. Assume first that $0 < a_1 < 1$. According to (7), the left-hand side of (10) is equal to

$$\left|\log|1-q_z| + \log|1-q_\tau/q_z| + \sum_{n=1}^{\infty}\log|1-q_\tau^n q_z| + \sum_{n=2}^{\infty}\log|1-q_\tau^n/q_z|\right|.$$

Since $0 < a_1 < 1$, both $|q_z|$ and $|q_\tau/q_z|$ are bounded by $|q_\tau|^{1/N}$, which does not exceed 0.1 because $|q_\tau| \le 10^{-N}$. (One can say even more: one of these numbers is bounded by $|q_\tau|^{1/N}$ and the other by $|q_\tau|^{1/2}$, which will be used later.) Hence, using (4) with $r = 0.1$, we obtain

$$\Bigl|\log|1-q_z| + \log|1-q_\tau/q_z|\Bigr| \le 2\cdot 1.1|q_\tau|^{1/N}. \tag{12}$$

Similarly, each of $|q_\tau^n q_z|$ and $|q_\tau^n/q_z|$ does not exceed 0.1, whence

$$\left|\sum_{n=1}^{\infty}\log|1-q_\tau^n q_z| + \sum_{n=2}^{\infty}\log|1-q_\tau^n/q_z|\right| \le 1.1\frac{|q_\tau q_z| + |q_\tau^2/q_z|}{1-|q_\tau|} \le 3|q_\tau|\cdot|q_\tau|^{1/N}, \tag{13}$$

which does not exceed $0.3|q_\tau|^{1/N}$. This proves (10).

When $a_1 = 0$ then $a_2 \notin \mathbb{Z}$ and $\zeta = e(a_2) \ne 1$. Further, we have $q_z = \zeta$ and the left-hand side of (10) is $\bigl|\sum_{n=1}^{\infty}\log|1-q_\tau^n\zeta| + \sum_{n=1}^{\infty}\log|1-q_\tau^n/\zeta|\bigr|$. Estimating the sums using (4), we obtain (11). $\square$

Since Siegel's functions has no poles nor zeros on the upper half plane $\mathcal{H}$, it should be bounded from above and from below on any compact subset of $\mathcal{H}$. In particular, it should be bounded where j is bounded. Here is a quantitative version of this.

Proposition 2.4 *Let $\mathbf{a} \in \mathbb{Q}^2$ be of order $N > 1$ in $(\mathbb{Q}/\mathbb{Z})^2$. Then for any $\tau \in \mathcal{H}$ we have*

$$\bigl|\log|g_{\mathbf{a}}(\tau)|\bigr| \le \frac{1}{12}\log\bigl(|j(\tau)| + 2200\bigr) + \log N + 0.1. \tag{14}$$

Proof Replacing τ by $\gamma\tau$ and $g_{\mathbf{a}}$ by $g_{\mathbf{a}\gamma^{-1}}$ with a suitable $\gamma \in \Gamma(1)$, we may assume that $\tau \in D$, and in particular $|q_\tau| < e^{-\pi\sqrt{3}}$. We may also assume that $0 \le a_1 < 1$. Now we argue as in the previous proof, the only difference being that in the case $0 < a_1 < 1$, we replace (12) by

$$\bigl|\log|1-q_z| + \log|1-q_\tau/q_z|\bigr| \le \Bigl|\log\bigl|1-e^{-\pi\sqrt{3}/N}\bigr|\Bigr| + \Bigl|\log\bigl|1-e^{-\pi\sqrt{3}/2}\bigr|\Bigr| \le \log N + 0.07. \quad (15)$$

Indeed, among the numbers $|q_z|$ and $|q_\tau/q_z|$ one is bounded by $|q_\tau|^{1/N}$ and the other is bounded by $|q_\tau|^{1/2}$, which implies (15).

Estimate (13) holds again, the right-hand side being bounded by $3|q_\tau| \le 0.02$. Thus, in the case $0 < a_1 < 1$ we have

$$\bigl|\log|g_{\mathbf{a}}(\tau)| - \ell_{\mathbf{a}} \log|q_\tau|\bigr| \le \log N + 0.1. \quad (16)$$

In the case $a_1 = 0$ we can use estimate (11), which implies $\bigl|\log|g_{\mathbf{a}}(\tau)| - \ell_{\mathbf{a}}\log|q_\tau|\bigr| \le \log 2 + 0.02$, and *a fortiori* (16).

Finally, Proposition 2.1 implies that $\bigl|\log|q_\tau|\bigr| \le \log\bigl(|j(\tau)| + 2200\bigr)$. Combining this (16), and using the inequality $|\ell_{\mathbf{a}}| \le 1/12$, we obtain (14). $\square$

2.3 Non-archimedean versions

We also need non-archimedean versions of some of the above inequalities. In this subsection K_v is a field complete with respect to a non-archimedean valuation v and $\bar{K}_v$ its algebraic closure. Let $q \in K_v$ satisfy $|q|_v < 1$. Put $j(q) = q^{-1} + 744 + 196884q + \ldots$ Further, for $\mathbf{a} = (a_1, a_2) \in \mathbb{Q}^2$ such that $\mathbf{a} \notin \mathbb{Z}^2$ put $q_z = q^{a_1}e(a_2)$ and define

$$g_{\mathbf{a}} = g_{\mathbf{a}}(q) = -q^{B_2(a_1)/2} e\left(\frac{a_2(a_1-1)}{2}\right)(1-q_z)\prod_{n=1}^{\infty}(1-q^n q_z)(1-q^n/q_z).$$

This expression is not well-defined because we use rational powers of q. However, if we fix $q^{1/2N^2} \in \bar{K}_v$, where N is the order of $\mathbf{a}$ in $(\mathbb{Q}/\mathbb{Z})^2$, then everything becomes well-defined, and, moreover, we again have (8) and (9). The statement of the following proposition is independent on the particular choice of $q^{1/2N^2}$. Recall that $\ell_{\mathbf{a}} = B_2\bigl(a_1 - \lfloor a_1 \rfloor\bigr)/2$.

Proposition 2.5 *In the above set-up, when $a_1 \notin \mathbb{Z}$ we have $\log|g_{\mathbf{a}}(q)|_v = \ell_{\mathbf{a}}\log|q|_v$, and when $a_1 \in \mathbb{Z}$ we have $\log|g_{\mathbf{a}}(q)|_v = \ell_{\mathbf{a}}\log|q|_v + \log|1-e(a_2)|_v$.*

Proof This is obvious when $0 \le a_1 < 1$, and the general case reduces to this one using (9). $\square$

Corollary 2.6 *In the above set-up, we have*

$$|\log|g_{\mathbf{a}}(q)|_v| \le \frac{1}{12}\log|j(q)|_v + \begin{cases} 0 & \textit{if } v(N) = 0 \\ \frac{\log p}{p-1} & \textit{if } v(N) > 0 \textit{ and } p \textit{ is the prime below } v. \end{cases}$$

Proof It suffices to notice that $|j(q)|_v = |q|_v^{-1}$, that $|\ell_{\mathbf{a}}| \le 1/12$, that $|1-e(a_2)|_v = 1$ if $v(N) = 0$, and that $1 \ge |1-e(a_2)|_v \ge p^{-1/(p-1)}$ if $v \mid p \mid N$. $\square$

3 Locating the "nearest cusp"

Let N be a positive integer, G a subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ and X_G the corresponding modular curve, defined over a number field K. In this section we fix a valuation v of K. We denote by $\mathcal{O}$ the ring of integers of K and by K_v the v-completion of K. When v is non-archimedean, we denote by $\mathcal{O}_v$ the ring of integers of K_v, and by k_v its residue field at v. As usual ζ_N will denote a primitive N-th root of unity. Recall that when we say that a curve "is defined over" a field, it means that this curve has a geometrically integral model over that field.

Let P be a point on $X_G(K_v)$ such that $|j(P)|$ is "large". Then it is intuitively clear that, in the v-adic metric, P is situated "near" a cusp of X_G. The purpose of this section is to make this intuitive observation precise and explicit. We shall locate this "nearest" cusp and specify what the word "near" means.

We first recall the following description of the cuspidal locus of $X(N)$ (for more details see e.g. [5, Chapitres V and VII)]). The cusps of $X(N)$ define a closed subscheme of the smooth locus of the modular model of $X(N)$ over $\mathbb{Z}[\zeta_N]$. Fix a uniformization $X(N)(\mathbb{C}) \simeq \bar{\mathcal{H}}/\Gamma(N)$, let c_∞ be the cusp corresponding to $\infty \in \bar{\mathcal{H}}$, and write $q^{1/N} = e^{2i\pi\tau/N}$ the usual parameter. If $c = \gamma(c_\infty)$, for some $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, is another cusp, denote by $q_c := q \circ \gamma^{-1}$ the parameter on $X(N)(\mathbb{C})$ at c. It follows from [5, Chapitre VII, Corollaire 2.5] that the completion of the curve $X(N)$ over $\mathbb{Z}[\zeta_N]$ along the section c is isomorphic to $\mathrm{Spec}\big(\mathbb{Z}[\zeta_N]\,[[q_c^{1/N}]]\big)$. In other words, the parameter $q_c^{1/N}$ at c on $X(N)(\mathbb{C})$ is actually defined over $\mathbb{Z}[\zeta_N]$, that is $q_c^{1/N}$ comes from an element of the completed local ring $\hat{\mathcal{O}}_{X(N),c}$ of the modular model of $X(N)$ over $\mathbb{Z}[\zeta_N]$, along the section c. Moreover the modular interpretation associates with each cusp a Néron polygon C with N sides on $\mathbb{Z}[\zeta_N]$, endowed with its structure of generalized elliptic curve, and enhanced with a basis of $C[N] \simeq \mathbb{Z}/N\mathbb{Z} \times \mu_N = \langle q^{1/N}, \zeta_N \rangle$ such that the determinant of this basis is 1, and two bases are identified if they are conjugate by the subgroup $\pm U = \pm \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$ of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$, the action being $\left(\begin{smallmatrix} \epsilon & a \\ 0 & \epsilon \end{smallmatrix}\right) : (q^{1/N}, \zeta_N) \mapsto (q^{\epsilon/N}\zeta_N^a, \zeta_N^\epsilon)$, where $\epsilon = \pm 1$ and $a \in \mathbb{Z}/N\mathbb{Z}$. We may, for instance, interpret c_∞ as the orbit $\big\{(C, (q^{\epsilon/N}\zeta_N^a, \zeta_N^\epsilon)),\ \epsilon \in \{\pm 1\},\ a \in \mathbb{Z}/N\mathbb{Z}\big\}$ of enhanced Néron polygons over $\mathbb{Z}[\zeta_N]$.

Next we describe the cusps on an arbitrary X_G. For each cusp c of X_G we obtain a parameter at c on X_G by picking a lift $\tilde{c}$ of c on $X(N)$ and taking the norm $\prod q_{\tilde{c}}^{1/N} \circ \gamma$, where γ runs through a set of representatives of $\Gamma/\Gamma(N)$ (recall that $\Gamma := \{\gamma \in \mathrm{SL}_2(\mathbb{Z}),\ (\gamma \mod N) \in G\}$). We denote by t_c this parameter in the sequel. Note that it is defined over a (possibly strict) subring of $\mathbb{Z}[\zeta_N]$. The modular interpretation of X_G associates to each cusp an orbit of our enhanced Néron polygon $\big(C, (q^{1/N}, \zeta_N)\big)$ under the action of the group generated by G and $\pm U$ given by $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) :\ \big(C, (q^{1/N}, \zeta_N)\big) \mapsto \big(C, (q^{a/N}\zeta_N^b, q^{c/N}\zeta_N^d)\big)$. It follows from the above that the cusps of X_G have values in a subring of $\mathbb{Z}[\zeta_N]$. Moreover, assume that X_G is defined over K, of which v is a place of characteristic p, with $N = p^n N'$ and $p \nmid N'$. Extending v to a place of $\mathcal{O}_v[\zeta_{N'}]$ if necessary, and setting $\mathcal{O}'_v := (\mathcal{O}[\zeta_{N'}])_v$, one sees that the closed subscheme of cusps over $\mathcal{O}'_v$ may be written as a sum of connected components of shape $\mathrm{Spec}(R)$ where R is a subring of $\mathcal{O}'_v[\zeta_{p^n}]$. Therefore if $v(N) = 0$, the subscheme of cusps is étale over $\mathcal{O}_v$, but this may not be the case if $v(N) > 0$. In the latter case, however, the ramification is well controlled. Indeed, with the preceding notations, set $\pi := (1 - \zeta_{p^n})$. Any two different p^n-th roots of unity $\zeta_{p^n}^a$ and $\zeta_{p^n}^b$ satisfy $(\zeta_{p^n}^a - \zeta_{p^n}^b) = \pi^{p^k}\alpha$ with α a v-invertible element and $0 \leq k \leq n-1$. It follows that Néron polygons enhanced with a level-N structure are distinct over $\mathbb{Z}[\zeta_N]/(\pi^{p^{n-1}+1})$. The modular interpretation shows more precisely that if two different cusps c_1 and c_2 have same reduction at v, then $t_{c_1}(c_2)$ has v-adic valuation less or equal to $1/(p-1)$ (if we normalize v to have $v(p) = 1$). This remark will be used later on.

To illustrate all this with a familiar example, letting $G := \left(\begin{smallmatrix} 1 & * \\ 0 & * \end{smallmatrix}\right) \subset \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$, which gives rise to the modular curve $X_1(N)$, one finds that there are $|(\mathbb{Z}/N\mathbb{Z})^\times|$ cusps, with modular interpretation corresponding to $\{(C, \zeta_N^{\epsilon a}) : \epsilon \in \{\pm 1\}\}$ where a runs through $(\mathbb{Z}/N\mathbb{Z})^\times/\pm 1$, and $\{(C, q^{\epsilon a/N}\zeta_N^\alpha) : \epsilon \in \{\pm 1\},\ \alpha \in (\mathbb{Z}/N\mathbb{Z})\}$, where a runs through the same set. The curve $X_1(N)$ is defined over $\mathbb{Q}$ and has a modular model over $\mathbb{Z}$. The cusps in the former subset above have values in $\mathbb{Z}\left[\zeta_N + \zeta_N^{-1}\right]$, and the cusps in the latter subset have values in $\mathbb{Z}$. In other words, the closed subscheme of cusps over $\mathbb{Z}$ is isomorphic to the disjoint union of $\mathrm{Spec}\left(\mathbb{Z}\left[\zeta_N + \zeta_N^{-1}\right]\right)$ and $|(\mathbb{Z}/N\mathbb{Z})^\times|/2$ copies of $\mathrm{Spec}(\mathbb{Z})$.

It is clear from the definition that the above parameter t_c defines a v-analytic function on a v-adic neighborhood of $c \in X_G(K_v)$ which satisfies the initial condition $t_c(c) = 0$. Further, if e_c is the ramification index of the covering $X_G \to X(1)$ at c (clearly, e_c divides N) then, setting $q_c := t_c^{e_c}$, the familiar expansion $j = q_c{}^{-1} + 744 + 196884 q_c + \ldots$ holds in a v-adic neighborhood of c, the right-hand side converging v-adically.

To be precise, t_c and q_c are defined and analytic on the set $\Omega_c = \Omega_{c,v}$ defined as follows. If v is

archimedean then $\Omega_c = Y_G(\bar{K}_v) \cup \{c\}$; in other words, Ω_c is $X_G(\bar{K}_v)$ with all the cusps except c taken away. If v is non-archimedean, then Ω_c consists of the points from $X_G(\bar{K}_v)$ having reduction c at v. Notice that $X_G(\bar{K}_v) = \bigcup_{c\in\mathcal{C}} \Omega_c$ if $v \in M_K^\infty$, and $\{P \in X_G(\bar{K}_v) : |j(P)|_v > 1\} = \bigcup_{c\in\mathcal{C}} \Omega_c$ if $v \in M_K^0$. More generally, since $|j(q_c)| = |q_c|_v^{-1}$ for a non-archimedean v, for any $R \geq 1$ and any $v \in M_K^0$ we have

$$\{P \in X_G(\bar{K}_v) : |j(P)|_v > R\} = \bigcup_{c\in\mathcal{C}} \{P \in \Omega_c : |q_c(P)|_v < R^{-1}\}, \tag{17}$$

which will be used later.

If v is non-archimedean and $v(N) = 0$, the sets Ω_c are pairwise disjoint, as in this case the cusps define a finite étale scheme over $\mathcal{O}_v$. In general however the sets Ω_c are not disjoint, so we need to refine them in order to be able to define the notion of "v-nearest cusp". Put

$$R_v = \begin{cases} 2500 & \text{if } v \in M_K^\infty, \\ 1 & \text{if } v \in M_K^0 \text{ and } v(N) = 0, \\ p^{N/(p-1)} & \text{if } v \in M_K^0 \text{ and } v \mid p \mid N, \end{cases} \qquad r_v = \begin{cases} 0.001 & \text{if } v \in M_K^\infty, \\ R_v^{-1} & \text{if } v \in M_K^0. \end{cases}$$

Finally, put

$$X_G(\bar{K}_v)^+ = \{P \in X_G(\bar{K}_v) : |j(P)|_v > R_v\}, \qquad \Omega_c^+ = \Omega_{c,v}^+ = \{P \in \Omega_c : |q_c(P)|_v < r_v\}.$$

Notice that $\Omega_c^+ = \Omega_c$ if v is non-archimedean and $v(N) = 0$.

Proposition 3.1 *In the above set-up, the sets Ω_c^+ are pairwise disjoint and we have*

$$X_G(\bar{K}_v)^+ \subseteq \bigcup_{c\in\mathcal{C}} \Omega_c^+ \tag{18}$$

with equality for the non-archimedean v.

The proposition implies that for every $P \in X_G(\bar{K}_v)^+$ there exists a unique cusp c such that $P \in \Omega_c^+$. We call it *the v-nearest cusp* (or simply the nearest cusp) to P.

Proof Assume first that v is archimedean, so that $\bar{K}_v = \mathbb{C}$. As above, let Γ be the pull-back of $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}) \cap G$ to $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$, and D be the usual fundamental domain for $\Gamma(1)$. Notice that $\Gamma(1)$ acts naturally on the set $\mathcal{C}$ of cusps, and that $q_{\gamma(c)} = q_c \circ \gamma^{-1}$ for $\gamma \in \Gamma(1)$, which implies that $\gamma(\Omega_c^+) = \Omega_{\gamma(c)}^+$.

Fix $P \in X_G(\mathbb{C}) = \bar{\mathcal{H}}/\Gamma$, and pick a representative $\tau \in \bar{\mathcal{H}}$ for P. Then there exists $\gamma \in \Gamma(1)$ such that $\gamma(\tau) \in \widetilde{D}$, where we put $\widetilde{D} = D \cup \{i\infty\}$. With the common abuse of notation, we denote by j both the j-invariant on $\mathcal{H}$ and on X_G, so that $j(\gamma(\tau)) = j(\tau) = j(P)$. Now if $P \in X_G(\mathbb{C})^+$ then $|j(\gamma(\tau))| = |j(P)| > 2500$, and Corollary 2.2 implies that $P^\gamma \in \Omega_{c_\infty}^+$. Hence $P \in \Omega_c^+$, where $c = \gamma^{-1}(c_\infty)$. This proves (18).

Now let us prove that the sets Ω_c^+ are pairwise disjoint. For $\tau \in \mathcal{H}$ the condition $|q_\tau| < 0.001$ implies that $\mathrm{Im}\,\tau > 1$. Hence

$$\{\tau \in \bar{\mathcal{H}} : |q_\tau| < 0.001\} \subset \bigcup_{\substack{\gamma\in\Gamma(1) \\ \gamma(i\infty)=i\infty}} \gamma\left(\widetilde{D}\right).$$

It follows that the pull-back of Ω_c^+ to $\bar{\mathcal{H}}$ is contained in the set $\Delta_c = \bigcup_{\substack{\gamma\in\Gamma(1) \\ \gamma(c_\infty)=c}} \gamma\left(\widetilde{D}\right)$. By the definition of $\widetilde{D}$, for $\gamma \neq \pm 1$ we have $\gamma(\widetilde{D}) \cap \widetilde{D} = \{i\infty\}$ if $\gamma(i\infty) = i\infty$, and $\gamma(\widetilde{D}) \cap \widetilde{D} = \varnothing$ otherwise. It follows that the sets Δ_c are pairwise disjoint. Hence so are the sets Ω_c^+. This completes the proof for archimedean v.

We now assume that v is non-archimedean. In this case (18) holds, with equality, as a particular case of (17), and we only need to show that the sets Ω_c^+ are pairwise disjoint. If $v(N)=0$ then, as already mentioned, the cusps of X_G define a finite étale closed subscheme of X_G over $\mathcal{O}_v$, so the sets $\Omega_c=\Omega_c^+$ are obviously pairwise disjoint.

Now assume $v(N)>0$. Let p be the residue characteristic of v and $p^n \| N$ be the largest power of p dividing $N=p^nN'$. As we have seen, the scheme of cusps on X_G may be no longer étale over $\mathcal{O}_v$. We can however still partition it into connected components, which totally ramify in the fiber at v. More precisely, setting as above ${\mathcal{O}'}_v:=(\mathcal{O}[\zeta_{N'}])_v$, each connected component over $\mathcal{O}'_v$ is schematically a $\mathrm{Spec}(R)$ for R a subring of ${\mathcal{O}'}_v[\zeta_{p^n}]$. Each set Ω_c contains exactly one such connected component of cusps, so when R does ramify nontrivially at v, then Ω_c is clearly "too large" (one has $\Omega_{c_1}=\Omega_{c_2}$ exactly when c_1 and c_2 have same reduction at v). We want to show that, nevertheless, the refined sets Ω_c^+ are pairwise disjoint.

If the cusps c_1 and c_2 belong to distinct connected components, then already Ω_{c_1} and Ω_{c_2} are disjoint, so $\Omega_{c_1}^+$ and $\Omega_{c_2}^+$ are disjoint *a fortiori*. Now assume that c_1 and c_2 belong to the same component, i.e. have same reduction at v. In this case, as explained before the proposition, one may write $t_{c_1}(c_2)=\pi^{p^k}a\in\mathcal{O}'_v[\zeta_{p^n}]$, for π a certain uniformizer (e.g. $\pi:=(\zeta_{p^n}-1)$), where the element a is v-invertible and $0\le k\le n-1$. As $v(\pi)=1/p^{n-1}(p-1)$ and $\Omega_{c_1}^+$ is contained in $\left\{P\in X_G(\bar{K}_v):|t_{c_1}(P)|_v<p^{-1/(p-1)}\right\}$, we see that c_2 does not belong to $\Omega_{c_1}^+$, which implies that the sets $\Omega_{c_1}^+$ and $\Omega_{c_2}^+$ are disjoint. This completes the proof of the proposition. □

Now Propositions 2.1 has the following consequence.

Proposition 3.2 *If v is archimedean then for $P\in X_G(K_v)^+$ with the nearest cusp c we have $\left|j(P)-q_c(P)^{-1}-744\right|_v\le 330000|q_c(P)|_v$. In particular, $\left|j(P)-q_c(P)^{-1}\right|_v\le 1100$ and*

$$\frac{3}{2}|j(P)|_v\ge\left|q_c(P)^{-1}\right|_v\ge\frac{1}{2}|j(P)|_v. \tag{19}$$

4 Modular Units

In this section we recall the construction of modular units on the modular curve X_G. By a *modular unit* we mean a rational function on X_G having poles and zeros only at the cusps.

4.1 Integrality of Siegel's Function

For $\mathbf{a}\in\mathbb{Q}^2\setminus\mathbb{Z}^2$ Siegel's function $g_{\mathbf{a}}$ (see Subsection 2.2) is algebraic over the field $\mathbb{C}(j)$: this follows from the fact that $g_{\mathbf{a}}^{12}$ is automorphic of level $2N^2$ [7, page 29]. Since $g_{\mathbf{a}}$ is holomorphic and does not vanish on the upper half-plane $\mathcal{H}$, both $g_{\mathbf{a}}$ and $g_{\mathbf{a}}^{-1}$ must be integral over the ring $\mathbb{C}[j]$. Actually, a stronger assertion holds.

Proposition 4.1 *Both $g_{\mathbf{a}}$ and $(1-\zeta_N)\,g_{\mathbf{a}}^{-1}$ are integral over $\mathbb{Z}[j]$. Here N is the exact order of* $\mathbf{a}$ *in $(\mathbb{Q}/\mathbb{Z})^2$ and ζ_N is a primitive N-th root of unity.*

This is, essentially, established in [7], but is not stated explicitly therein. Therefore we briefly indicate the proof here. Recall that a holomorphic and $\Gamma(N)$-automorphic function $f:\mathcal{H}\to\mathbb{C}$ admits the *infinite q-expansion*

$$f(\tau)=\sum_{k\in\mathbb{Z}}a_kq^{k/N}, \tag{20}$$

where $q=q_\tau=e^{2\pi i\tau}$. We call the q-series (20) *algebraic integral* if the following two conditions are satisfied: the negative part of (20) has only finitely many terms (that is, $a_k=0$ for large negative k), and the coefficients a_k are algebraic integers. Algebraic integral q-series form a ring. The invertible elements of this ring are q-series with invertible leading coefficient. By the *leading coefficient* of an algebraic integral q-series we mean a_m, where $m\in\mathbb{Z}$ is defined by $a_m\ne 0$, but $a_k=0$ for $k<m$.

Lemma 4.2 *Let f be a $\Gamma(N)$-automorphic function such that for every $\gamma\in\Gamma(1)$ the q-expansion of $f\circ\gamma$ is algebraic integral. Then f is integral over $\mathbb{Z}[j]$.*

Proof This is, essentially, Lemma 2.1 from [7, Section 2.2]. Since f is $\Gamma(N)$-automorphic, the set $\{f\circ\gamma : \gamma\in\Gamma(1)\}$ is finite. The coefficients of the polynomial $F(T)=\prod(T-f\circ\gamma)$ (where the product is taken over the finite set above) are $\Gamma(1)$-automorphic functions with algebraic integral q-expansions. By the q-expansion principle, the coefficients of $F(T)$ belong to $\bar{\mathbb{Z}}[j]$, where $\bar{\mathbb{Z}}$ is the ring of all algebraic integers. It follows that f is integral over $\bar{\mathbb{Z}}[j]$, hence over $\mathbb{Z}[j]$. □

Proof of Proposition 4.1 The function $g_{\mathbf{a}}^{12}$ is automorphic of level $2N^2$ and its q-expansion is algebraic integral (as one can easily see by transforming the infinite product (7) into an infinite series). By (8), the same is true for for every $(g_{\mathbf{a}}\circ\gamma)^{12}$. Lemma 4.2 now implies that $g_{\mathbf{a}}^{12}$ is integral over $\mathbb{Z}[j]$, and so is $g_{\mathbf{a}}$.

Further, the q-expansion of $g_{\mathbf{a}}$ is invertible if $a_1\notin\mathbb{Z}$ and is $1-e(a_2)$ times an invertible q-series if $a_1\in\mathbb{Z}$. Hence the q-expansion of $g_{\mathbf{a}}^{-1}$ is algebraic integral when $a_1\notin\mathbb{Z}$, and if $a_1\in\mathbb{Z}$ the same is true for $(1-e(a_2))\,g_{\mathbf{a}}^{-1}$. In the latter case N is the exact order of a_2 in $\mathbb{Q}/\mathbb{Z}$, which implies that $(1-\zeta_N)/\,(1-e(a_2))$ is an algebraic unit. Hence, in any case, $(1-\zeta_N)g_{\mathbf{a}}^{-1}$ has algebraic integral q-expansion, and the same is true with $g_{\mathbf{a}}$ replaced by $g_{\mathbf{a}}\circ\gamma$ for any $\gamma\in\Gamma(1)$ (we again use (8) and notice that $\mathbf{a}$ and $\mathbf{a}\gamma$ have the same order in $(\mathbb{Q}/\mathbb{Z})^2$). Applying Lemma 4.2 to the function $\big((1-\zeta_N)g_{\mathbf{a}}^{-1}\big)^{12}$, we complete the proof. □

4.2 Modular Units on $X(N)$

From now on, we fix an integer $N>1$. Recall that the curve $X(N)$ is defined over the field $\mathbb{Q}(\zeta_N)$. Moreover, the field $\mathbb{Q}\big(X(N)\big)=\mathbb{Q}(\zeta_N)\big(X(N)\big)$ is a Galois extension of $\mathbb{Q}(j)$, the Galois group being isomorphic to $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$. The isomorphism

$$\mathrm{Gal}\left(\mathbb{Q}\big(X(N)\big)/\,\mathbb{Q}(j)\right)\cong\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z}) \tag{21}$$

is defined up to an inner automorphism; once it is fixed, we have the well-defined isomorphisms

$$\mathrm{Gal}\left(\mathbb{Q}\big(X(N)\big)/\,\mathbb{Q}(\zeta_N,j)\right)\cong\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}),\qquad \mathrm{Gal}\big(\mathbb{Q}(\zeta_N)/\,\mathbb{Q}\big)\cong(\mathbb{Z}/N\mathbb{Z})^\times, \tag{22}$$

and we may identify the groups on the left and on the right in (21 and 22). Our choice of the isomorphism (21) will be specified in Proposition 4.3.

According to Theorem 1.2 from [7, Section 2.1], given $\mathbf{a}=(a_1,a_2)\in(N^{-1}\mathbb{Z})^2\setminus\mathbb{Z}^2$, the function $g_{\mathbf{a}}^{12N}$ is $\Gamma(N)$-automorphic of weight 0. Hence $g_{\mathbf{a}}^{12N}$ defines a rational function on the modular curve $X(N)$, to be denoted by $u_{\mathbf{a}}$. Since the root of unity in (9) is of order dividing $12N$, we have $u_{\mathbf{a}}=u_{\mathbf{a}'}$ when $\mathbf{a}\equiv\mathbf{a}' \mod \mathbb{Z}^2$. Hence $u_{\mathbf{a}}$ is well-defined when $\mathbf{a}$ is a non-zero element of the abelian group $(N^{-1}\mathbb{Z}/\mathbb{Z})^2$, which will be assumed in the sequel. We put $\mathbf{A}=(N^{-1}\mathbb{Z}/\mathbb{Z})^2\setminus\{0\}$.

The functions $u_{\mathbf{a}}$ have the following properties.

Proposition 4.3 *(a) The functions $u_{\mathbf{a}}$ and $(1-\zeta_{N_{\mathbf{a}}})^{12N}u_{\mathbf{a}}^{-1}$ are integral over $\mathbb{Z}[j]$, where $N_{\mathbf{a}}$ is the exact order of $\mathbf{a}$ in $(N^{-1}\mathbb{Z}/\mathbb{Z})^2$. In particular, $u_{\mathbf{a}}$ has zeros and poles only at the cusps of $X(N)$.*

(b) The functions $u_{\mathbf{a}}$ belong to the field $\mathbb{Q}\big(X(N)\big)$, and the Galois action on the set $\{u_{\mathbf{a}}\}$ over $\mathbb{Q}(j)$ is compatible with the (right) linear action of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ on $\mathbf{A}$ in the following sense: the isomorphism (21) can be chosen so that for any $\sigma\in\mathrm{Gal}\Big(\mathbb{Q}\big(X(N)\big)/\,\mathbb{Q}(j)\Big)=\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ and any $\mathbf{a}\in\mathbf{A}$ we have $u_{\mathbf{a}}^\sigma=u_{\mathbf{a}\sigma}$.

(c) For the cusp c_∞ at infinity we have $\mathrm{ord}_{c_\infty}u_{\mathbf{a}}=12N^2\ell_{\mathbf{a}}$, where $\ell_{\mathbf{a}}$ is defined in Subsection 1.1. For an arbitrary cusp c we have $|\mathrm{ord}_c u_{\mathbf{a}}|\le N^2$.

Proof Item (a) follows from Proposition 4.1. Item (b) is Proposition 1.3 from [7, Chapter 2]. We are left with item (c). The order of vanishing of $u_{\mathbf{a}}$ at $i\infty$ is $12N\ell_{\mathbf{a}}$. Since the ramification index of the covering $X(N)\to X(1)$ at every cusp is N, we obtain $\mathrm{ord}_{c_\infty}u_{\mathbf{a}}=12N^2\ell_{\mathbf{a}}$. Since $|\ell_{\mathbf{a}}|\le 1/12$,

we have $|\mathrm{ord}_{c_\infty} u_{\mathbf{a}}| \le N^2$. The case of arbitrary c reduces to the case $c = c_\infty$ upon replacing $\mathbf{a}$ by $\mathbf{a}\sigma$ where $\sigma \in \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$ is such that $\sigma(c) = c_\infty$. ☐

The group generated by the principal divisors $(u_{\mathbf{a}})$, where $\mathbf{a} \in \mathbf{A}$, is contained in the group of cuspidal divisors on $X(N)$ (that is, the divisors supported at the set $\mathcal{C}(N) = \mathcal{C}(\Gamma(N))$ of cusps). Since principal divisors are of degree 0, the rank of the former group is at most $|\mathcal{C}(N)| - 1$. It is fundamental for us that this rank is indeed maximal possible. The following proposition is Theorem 3.1 in [7, Chapter 2].

Proposition 4.4 *The group generated by the set $\{(u_{\mathbf{a}}) : \mathbf{a} \in \mathbf{A}\}$ is of rank $|\mathcal{C}(N)| - 1$.* ☐

We also need to know the behavior of the functions $u_{\mathbf{a}}$ near the cusps, and estimate them in terms of the modular invariant j. In the following proposition K is a number field containing ζ_N and v is a valuation of K, extended somehow to $\bar{K}$. We use the notation of Section 3.

Proposition 4.5 *(a) Let c be a cusp of $X(N)$. If $v \in M_K^\infty$ then*

$$\begin{aligned} \bigl|\log|u_{\mathbf{a}}(P)|_v - \mathrm{ord}_c u_{\mathbf{a}} \log|t_c(P)|_v\bigr| &\le 36N|q_c(P)|_v^{1/N} \quad \text{when } a_1 \ne 0,\\ \bigl|\log|u_{\mathbf{a}}(P)|_v - \mathrm{ord}_c u_{\mathbf{a}} \log|t_c(P)|_v - 12N\log|1-e(a_2)|_v\bigr| &\le 36N|q_c(P)|_v \qquad \text{when } a_1 = 0 \end{aligned}$$

for any $P \in \Omega_{c,v}$ such that $|q_c(P)|_v < 10^{-N}$. If $v \in M_K^0$ then

$$\log|u_{\mathbf{a}}(P)|_v = \begin{cases} \mathrm{ord}_c u_{\mathbf{a}} \log|t_c(P)|_v & \text{when } a_1 \ne 0,\\ \mathrm{ord}_c u_{\mathbf{a}} \log|t_c(P)|_v + 12N\log|1-e(a_2)|_v & \text{when } a_1 = 0 \end{cases}$$

for any $P \in \Omega_{c,v}$.

(b) If $v \in M_K^\infty$ then

$$\bigl|\log|u_{\mathbf{a}}(P)|_v\bigr| \le N\log\bigl(|j(P)|_v + 2200\bigr) + 14N\log N$$

for any $P \in X(N)(K_v)$. If $v \in M_K^0$ then $\bigl|\log|u_{\mathbf{a}}(P)|_v\bigr| \le N\log|j(P)|_v + \log R_v$ for any $P \in X(N)(K_v)$ such that $|j(P)|_v > 1$.

Proof When $c = c_\infty$ this is an immediate consequence of Propositions 2.3, 2.4, 2.5 (notice that $\log|q_c|v{=}N\log|t_c|_v$ for every cusp c) and Corollary 2.6. The general case reduces to the case $c = c_\infty$ by applying a suitable Galois automorphism. ☐

4.3 K-rational Modular Units on X_G

Now let K be a number field, and let G be a subgroup of $\mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$. Let $\det G$ be the image of G under the determinant map $\det\colon \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z}) \to (\mathbb{Z}/N\mathbb{Z})^\times = \mathrm{Gal}(\mathbb{Q}(\zeta_N)/\mathbb{Q})$ (recall that we have a well-defined isomorphism (22)). In the sequel we shall assume that $K \supseteq \mathbb{Q}(\zeta_N)^{\det G}$, where $\mathbb{Q}(\zeta_N)^{\det G}$ is the subfield of $\mathbb{Q}(\zeta_N)$ stable under $\det G$. This assumption implies that the curve X_G is defined over K. Then $G' := \mathrm{Gal}\left(K\bigl(X(N)\bigr)\big/ K\left(X_G\right)\right)$ is a subgroup of G. For every $\mathbf{a} \in \mathbf{A}$ we put $w_{\mathbf{a}} = \prod_{\sigma\in G'} u_{\mathbf{a}\sigma}$. Since $u_{\mathbf{a}\sigma} = u_{\mathbf{a}}^\sigma$, the functions $w_{\mathbf{a}}$ are contained in $K(X_G)$. They have the following properties.

Proposition 4.6 *(a) The functions $w_{\mathbf{a}}$ have zeros and poles only at the cusps of X_G. If c is such a cusp, then $|\mathrm{ord}_c w_{\mathbf{a}}| \le |G'|N^2$.*

(b) For every $\mathbf{a} \in \mathbf{A}$ there exists an algebraic integer $\lambda_{\mathbf{a}} \in \mathbb{Z}[\zeta_N]$, which is a product of $|G'|$ factors of the form $(1-\zeta_{N'})^{12N}$, where $N' \mid N$, such that the functions $w_{\mathbf{a}}$ and $\lambda_{\mathbf{a}} w_{\mathbf{a}}^{-1}$ are integral over $\mathbb{Z}[j]$.

(c) *If $v \in M_K^\infty$ then*

$$\left|\log|w_{\mathbf{a}}(P)|_v\right| \le |G'|N\log\bigl(|j(P)|_v + 2200\bigr) + 14|G'|N\log N$$

for any $P \in X_G(K_v)$. If $v \in M_K^0$ then

$$\left|\log|w_{\mathbf{a}}(P)|_v\right| \le |G'|N\log|j(P)|_v + |G'|\log R_v$$

for any $P \in X(N)(K_v)$ such that $|j(P)|_v > 1$.

(d) *For every $\mathbf{a} \in \mathbf{A}$ and every cusp c there exists an algebraic integer $\beta = \beta(\mathbf{a}, c) \in \mathbb{Z}[\zeta_N]$, which is a product of at most $|G'|$ factors of the form $(1 - e(a))^{12N}$, where $a \in N^{-1}\mathbb{Z}/\mathbb{Z}$ and $a \neq 0$, such that for any $v \in M_K$ and for any $P \in \Omega_{c,v}$ we have the following. If v is archimedean and $|q_c(P)|_v \le 10^{-N}$ then*

$$\left|\log|w_{\mathbf{a}}(P)|_v - \mathrm{ord}_c w_{\mathbf{a}}\log|t_c(P)|_v - \log|\beta|_v\right| \le 36|G'|N|q_c(P)|_v^{1/N}.$$

If v is non-archimedean then $\log|w_{\mathbf{a}}(P)|_v = \mathrm{ord}_c w_{\mathbf{a}}\log|t_c(P)|_v + \log|\beta|_v$.

(e) *The group generated by the principal divisors $(w_{\mathbf{a}})$ is of rank $|\mathcal{C}(G,K)| - 1$.*

Proof Items (a) and (b) follow from Proposition 4.3, items (c) and (d) follow from Proposition 4.5. Finally, item (e) follows from Proposition 4.4 and Lemma 4.7 below. On should apply the lemma (whose proof is left to the reader) with A as the group of degree 0 cuspidal divisors on $X(N)$, with B as the group of all cuspidal divisors on $X(N)$ generated by the principal divisors $(u_{\mathbf{a}})$, and with G as G'. □

Lemma 4.7 *Let G be a finite group, and let A be a torsion-free finitely generated (left) G-module. For $a \in A$ put $a_G = \sum_{g\in G} ga$ and denote by A^G the submodule of the G-invariant elements. Further, let B be a finite index subgroup of A. Then $B_G = \{b_G : b \in B\}$ is a finite index submodule of A^G.* □

4.4 A Unit Vanishing at the Given Cusps

Item (e) of Proposition 4.6 implies that for any proper subset of $\mathcal{C}(G,K)$ there is a K-rational unit on X_G vanishing at this subset. In this subsection we give a quantitative version of this fact. We shall use the following simple lemma, where we denote by $\|\cdot\|_1$ the ℓ_1-norm.

Lemma 4.8 *Let M be an $s \times t$ matrix of rank s with entries in $\mathbb{Z}$. Assume that the entries of M do not exceed A in absolute value. Then there exists a vector $\mathbf{b} \in \mathbb{Z}^t$ such that $\|\mathbf{b}\|_1 \le s^{s/2+1}A^{s-1}$, and such that all the s coordinates of the vector $M\mathbf{b}$ (in the standard basis) are strictly positive.*

Proof Assume first that $s = t$. Let d be the determinant of M. Then the column vector $(|d|, \dots, |d|)$ can be written as $M\mathbf{b}$, where $\mathbf{b} = (b_1, \dots, b_s)$ with b_k being (up to the sign) the determinant of the matrix obtained from M upon replacing the k-th row by $(1, \dots, 1)$. Using Hadamard's inequality, we bound $|b_k|$ by $\sqrt{s}\left(\sqrt{s}A\right)^{s-1}$. This proves the lemma in the case $s = t$. The general case reduces to the case $s = t$ by selecting a non-singular $s \times s$ sub-matrix, which gives s entries of the vector $\mathbf{b}$; the remaing $t - s$ entries are set to be 0. □

Now let G, K and G' be as in Subsection 4.3.

Proposition 4.9 *Let Σ be a proper subset of $\mathcal{C}(G,K)$. Assume that $|\Sigma| \le s$, and put $B = s^{s/2+1}\left(|G'|N^2\right)^{s-1}$. Then there exists a K-rational modular unit w on X_G with the following properties.*

(a) *If c is a cusp such that the orbit of c is in Σ then $\mathrm{ord}_c w > 0$.*

(b) For every cusp c we have

$$|\mathrm{ord}_c w| \le B|G'|N^2. \tag{23}$$

(c) There exists an algebraic integer λ, which is a product of at most $|G'|B$ factors of the form $(1-\zeta_{N'})^{12N}$, where $N' \mid N$, such that λw is integral over $\mathbb{Z}[j]$.

(d) If $v \in M_K^\infty$ then for any $P \in X_G(K_v)$ we have

$$\bigl|\log|w(P)|_v\bigr| \le B|G'|N\log\bigl(|j(P)|_v + 2200\bigr) + 14B|G'|N\log N.$$

If $v \in M_K^0$ then for any $P \in X(N)(K_v)$ such that $|j(P)|_v > 1$ we have

$$\bigl|\log|w(P)|_v\bigr| \le B|G'|N\log|j(P)|_v + B|G'|\log R_v.$$

(e) For every cusp c there exists an algebraic number $\beta = \beta(c) \in \mathbb{Z}[\zeta_N]$ which is a product of at most $|G'|B$ factors of the form $(1-e(a))^{\pm 12N}$, where $a \in N^{-1}\mathbb{Z}/\mathbb{Z}$ and $a \ne 0$, such that for any $v \in M_K$ and for any $P \in \Omega_{c,v}$ we have the following. If v is archimedean and $|q_c(P)|_v \le 10^{-N}$ then

$$\bigl|\log|w(P)|_v - \mathrm{ord}_c w \log|t_c(P)|_v - \log|\beta|_v\bigr| \le 36B|G'|N|q_c(P)|_v^{1/N}.$$

If v is non-archimedean then $\log|w(P)|_v = \mathrm{ord}_c w\log|t_c(P)|_v + \log|\beta|_v$.

Proof The K-rational Galois orbit of a cusp c has $[K(c):K]$ elements. Fix a representative in every such orbit and consider the $|\mathcal{C}(G,K)| \times |\mathbf{A}|$ matrix $(\mathrm{ord}_c w_{\mathbf{a}})$, where c runs over the set of selected representatives. According to item (e) of Proposition 4.6, this matrix is of rank $|\mathcal{C}(G,K)|-1$, and the only (up to proportionality) linear relation between the rows is $\sum_c [K(c):K]\mathrm{ord}_c w_{\mathbf{a}} = 0$ for every $\mathbf{a} \in \mathbf{A}$. It follows that any proper subset of the rows of our matrix is linearly independent. In particular, if we select the rows corresponding to the set Σ, we get a sub-matrix of rank $|\Sigma|$. Applying to it Lemma 4.8, where we may take $A = |G'|N^2$ due to item (a) of Proposition 4.6, we find integers $b_{\mathbf{a}}$ such that $\sum_{\mathbf{a}\in\mathbf{A}} |b_{\mathbf{a}}| \le B$ and such that the function $w = \prod_{\mathbf{a}\in\mathbf{A}} w_{\mathbf{a}}^{b_{\mathbf{a}}}$ is as wanted. □

5 Proof of Theorem 1.2

We use the notation of Section 3. We put $R_v' = 50^N$ for archimedean v and $R_v' = R_v$ for non-archimedean v. Since $N \ge 2$, we have $R_v' \ge R_v$.

We use the notation $d_v = [K_v : \mathbb{Q}_v]$ and $d = [K:\mathbb{Q}]$. We fix an extension of every $v \in M_K$ to $\bar{K}$ and denote this extension by v as well.

We shall use the estimate

$$\mathcal{R} \le \sum_{p\mid N} \frac{\log p}{p-1} \le \omega(N)\log 2 \le \log N, \tag{24}$$

for the quantity $\mathcal{R}$, defined in (3). (Here $\omega(N)$ is the number of prime divisors of N.) Of course, much sharper estimates for $\mathcal{R}$ are possible as well, but (24) is plainly sufficient for us.

5.1 The Runge Unit

Fix $P \in Y_G(\mathcal{O}_S)$. Let S_1 consist of the places $v \in M_K$ such that $|j(P)|_v > R_v'$. Plainly, $S_1 \subset S$. Since $R_v' \ge R_v$, Proposition 3.1 applies to our P and every $v \in S_1$. Thus, for $v \in S_1$ let c_v be the v-nearest cusp to P, and let Σ be the set of all $\mathrm{Gal}(\bar{K}/K)$-orbits of cusps containing some of the c_v. Then $|\Sigma| \le |S_1| \le |S|$, and since $|S| < |\mathcal{C}(G,K)|$ by the assumption, Σ is a proper subset of $\mathcal{C}(G,K)$. Let w and B be as in Proposition 4.9, where we may put $s = |S|$. Then $\mathrm{ord}_{c_v} w > 0$ for every $v \in S_1$, and the other statements of this proposition are satisfied.

Since w is a modular unit and P is not a cusp, we have $w(P) \ne 0, \infty$, and the product formula gives $\sum_{v\in M_K} d_v \log|w(P)|_v = 0$. We want to show that this is impossible when $\mathrm{h}(P)$ is too large.

5.2 Partitioning the Places of K

We partition the set of places M_K into three pairwise disjoint subsets: $M_K = S_1 \cup S_2 \cup S_3$, where $S_i \cap S_j = \varnothing$ for $i \neq j$. The set S_1 is already defined. The set S_2 consists of the *archimedean* places not belonging to S_1 and the *non-archimedean* places v not belonging to S_1 and such that $|j(P)|_v > 1$. (Obviously, $S_2 \subset S$.) Finally, the set S_3 consists of the places not belonging to $S_1 \cup S_2$; in other words, $v \in S_3$ if and only if v is non-archimedean and $|j(P)|_v \leq 1$.

We will estimate from above the three sums $\Xi_i = \sum_{v\in S_i} d_v \log|w(P)|_v$. We will show that $\Xi_1 \leq -N^{-1} d\mathrm{h}(P) + O(1)$, where the $O(1)$-term is independent of P (it will be made explicit). Further, we will bound Ξ_2 and Ξ_3 independently of P. Since

$$\Xi_1 + \Xi_2 + \Xi_3 = 0, \tag{25}$$

(which is a different writing of the product formula), this would bound $\mathrm{h}(P)$.

5.3 Estimating Ξ_1

For $v \in S_1$ we have $P \in \Omega_{c_v,v}$, we may apply item (e) of Proposition 4.9. Since $\mathrm{ord}_{c_v} w > 0$ and $\log q_{c_v}(P) = e \log t_{c_v}(P)$ with $e \mid N$, we have, for an archimedean $v \in S_1$

$$\begin{aligned}
\log|w(P)|_v &\leq \frac{\mathrm{ord}_{c_v} w}{e} \log|q_{c_v}(P)|_v + \log|\beta(c_v)|_v + 36B|G'|N|q_{c_v}(P)|_v^{1/N} \\
&\leq -\frac{\mathrm{ord}_{c_v} w}{N} \log|j(P)|_v + \log|\beta(c_v)|_v + 2B|G'|N \qquad \text{(we use (19) and (23))} \qquad (26)\\
&\leq -N^{-1} \log|j(P)|_v + \log|\beta(c_v)|_v + 2B|G'|N. \qquad (27)
\end{aligned}$$

For a non-archimedean $v \in S_1$ we have

$$\log|w(P)|_v \leq N^{-1} \log|q_c(P)|_v + \log|\beta(c)|_v = -N^{-1} \log|j(P)|_v + \log|\beta(c_v)|_v. \tag{28}$$

Next, we want to estimate $\sum_{v\in S_1} \log|\beta(c_v)|_v$. Recall that $\beta(c)$ is a product of at most $12B|G'|N$ numbers of the type $1 - e(a)$, where a is a non-zero element of $N^{-1}\mathbb{Z}/\mathbb{Z}$. For such a we have $1/N \leq \left|1 - e(a)\right|_v \leq 2$ if v is archimedean, $p^{-1/(p-1)} \leq \left|1 - e(a)\right|_v \leq 1$ if v is non-archimedean and $v \mid p \mid N$, and $\left|1 - e(a)\right|_v = 1$ if v is non-archimedean and $v(N) = 0$. It follows that

$$\sum_{v\in S_1} \log|\beta(c_v)|_v \leq 12dB|G'|N\left(\log N + \mathcal{R}\right) \leq 24dB|G'|N \log N, \tag{29}$$

where $\mathcal{R}$ is defined in (3) and is estimated using (24).

Thus, combining (27), (28) and (29), we obtain

$$\begin{aligned}
\Xi_1 &\leq -N^{-1} \sum_{v\in S_1} d_v \log|j(P)|_v + \sum_{v\in S_1} \log|\beta(c_v)|_v + 2dB|G'|N \\
&\leq -N^{-1} \sum_{v\in S_1} d_v \log|j(P)|_v + 27dB|G'|N \log N.
\end{aligned}$$

Further, since $|j(P)|_v \leq R'_v$ for $v \in S \setminus S_1$, we have

$$\sum_{v\in S\setminus S_1} d_v \log|j(P)|_v \leq \sum_{v\in M_K} d_v \log R'_v \leq dN\left(\log 50 + \mathcal{R}\right) \leq dN \log(50N),$$

by (24), and we obtain

$$\begin{aligned}
\Xi_1 &\leq -N^{-1} \sum_{v\in S} d_v \log|j(P)|_v + dN\big(27|G'|B \log N + \log(50N)\big) \\
&\leq -N^{-1} \sum_{v\in S} d_v \log|j(P)|_v + 32dB|G'|N \log N.
\end{aligned}$$

Finally, since $j(P)$ is an S-integer, we have

$$\mathrm{h}(P)=\mathrm{h}(j(P))=d^{-1}\sum_{v\in S}d_v\log^+|j(P)|_v\ge d^{-1}\sum_{v\in S}d_v\log|j(P)|_v,$$

and we obtain

$$\Xi_1\le d\left(-N^{-1}\mathrm{h}(P)+31B|G'|N\log N\right). \tag{30}$$

5.4 Estimating Ξ_2, Ξ_3 and Completing the Proof

Item (d) of Proposition 4.9 implies that for an archimedean $v\in S_2$

$$\log|w(P)|_v\le B|G'|N\left(\log\left(50^N+2200\right)+14\log N\right)\le 10B|G'|N^2,$$

and for a non-archimedean $v\in S_2$ we have $\left|\log|w(P)|_v\right|\le B|G'|N\log R_v+B|G'|\log R_v$. Using (24), we obtain

$$\Xi_2\le 10dB|G'|N^2+dB|G'|(N^2+N)\mathcal{R}\le dB|G'|N^2(\mathcal{R}+11). \tag{31}$$

Futher, let λ be from item (c) of Proposition 4.9. Then $\mathrm{h}(\lambda)\le 12B|G'|N\log 2\le 9B|G'|N$, because $\mathrm{h}(1-\zeta)\le\log 2$ for a root of unity ζ. For $v\in S_3$ the number $j(P)$ is a v-adic integer. Hence so is the number $\lambda w(P)$. It follows that $|w(P)|_v\le|\lambda^{-1}|_v$ for $v\in S_3$, and

$$\Xi_3\le\sum_{v\in S_3}d_v\log\left|\lambda^{-1}\right|_v\le d\mathrm{h}\left(\lambda^{-1}\right)=d\mathrm{h}(\lambda)\le 9dB|G'|N. \tag{32}$$

Combining this with (25), (30) and (31), we obtain $\mathrm{h}(P)\le B|G'|N^3(\mathcal{R}+30)$, which is (2) with $|G|$ replaced by $|G'|$. $\square$

6 Proof of Theorem 1.1

It is similar and simpler than that of Theorem 1.2. In this case $d=1$ and S consists of the infinite place of $\mathbb{Q}$, whence $s=1$, $B=1$ and $\mathcal{R}=0$. We may take as the Runge unit w one of the functions $w_{\mathbf{a}}^{\pm1}$. We may assume that $S_1=S$ and $S_2=\varnothing$; otherwise we would have the estimate $\log|j(P)|\le N\log 50$, which is much sharper than (1). We denote by c the nearest cusp to P with respect to the infinite place.

Thus, $\Xi_1=\log|w(P)|$, and (27) now reads $\log|w(P)|\le -N^{-1}\log|j(P)|+\log|\beta(c)|+2|G'|N$. Estimating $\left|\log|\beta(c)|\right|$ by $12|G'|N\log N$, we obtain $\Xi_1\le -N^{-1}\log|j(P)|+15|G'|N\log N$. Further, $\Xi_2=0$, and Ξ_3 can be estimated by $9|G'|N$, according to (32). Since $\Xi_1+\Xi_3=0$, we obtain (1) with $|G|$ replaced by $|G'|$. $\square$

7 A Special Case

When passing from (26) to (27), we used the trivial estimate $\mathrm{ord}_c w\ge 1$. One can improve our results by using a more elaborate lower bound for $\mathrm{ord}_c w$. In this section we apply this approach to the case when G is the normalizer of a split torus of prime level, which is important for the subsequent applications. Of course, similar strategy can be used in many different cases as well.

We start by recalling definitions and notations that will be in force for the rest of the article. Recall that a Cartan subgroup of the algebraic group GL_2 over some ring is a maximal subtorus, which can be either (totally) split or nonsplit. More precisely, fix a prime number p. The subgroup of diagonal matrices in $\mathrm{GL}_2(\mathbb{Z}_p)$ is a split Cartan subgroup, whose normalizer consists in diagonal and antidiagonal matrices. Given an integer $n\ge 0$, the normalizer of a split Cartan subgroup is the image in $\mathrm{GL}_2(\mathbb{Z}/p^n\mathbb{Z})$ of a group conjugate to the above subgroup of $\mathrm{GL}_2(\mathbb{Z}_p)$. If G is such a group mod p^n, we denote by $X_{\mathrm{split}}(p^n)$ the corresponding modular curve over $\mathbb{Q}$, and

by $Y_{\mathrm{split}}(p^n)$ its finite part. On the other hand, let $\mathbb{Z}_{p^2}$ be the ring of integers of the unramified quadratic extension of $\mathbb{Q}_p$. By making the group $\mathbb{Z}_{p^2}^\times$ act on $\mathbb{Z}_{p^2}$ by multiplication, the choice of a $\mathbb{Z}_p$-basis of $\mathbb{Z}_{p^2}$ defines an embedding of $\mathbb{Z}_{p^2}^\times$ into $\mathrm{GL}_2(\mathbb{Z}_p)$. The image of such an embedding is by definition a non split Cartan subgroup of $\mathrm{GL}_2(\mathbb{Z}_p)$; it has index 2 in its normalizer. For n any positive integer, reduction mod p^n similarly defines (normalizer of) nonsplit Cartan subgroups in $\mathrm{GL}_2(\mathbb{Z}/p^n\mathbb{Z})$. Those subgroups define in the usual way modular curves over $\mathbb{Q}$, which we denote by $X_{\mathrm{nonsplit}}(p^n)$.

Now we focus on the case where $N = p$ is an odd prime number, and we let G be the normalizer of a split Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$.

Theorem 7.1 *For any $P \in Y_{\mathrm{split}}(p)(\mathbb{Z})$ we have $\log|j(P)| \le 23p\log p$.*

(Note that Theorem 1.1 with $N = p$ gives the bound $\log|j(P)| \le 60p^2(p-1)^2\log p$.)

The curve $X_{\mathrm{split}}(p)$ has $(p+1)/2$ cusps, among which one (the cusp at infinity c_∞) is defined over $\mathbb{Q}$ and the other $(p-1)/2$ are conjugate over $\mathbb{Q}$, so that we have exactly 2 Galois orbits of cusps over $\mathbb{Q}$. The covering $X(p) \to X_{\mathrm{split}}(p)$ is unramified at the cusps, and the covering $X_{\mathrm{split}}(p) \to X(1)$ has ramification p at every cusp.

We shall use the following lemma.

Lemma 7.2 *Let G be the normalizer of a split Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$ and c a cusp of $X_G = X_{\mathrm{split}}(p)$. Then for any $\mathbf{a} \in \mathbf{A}$ we have*

$$|\mathrm{ord}_c w_{\mathbf{a}}| \ge 2p(p-1)^2 = p|G|. \tag{33}$$

Proof The proof relies on the identity $\sum_{k=1}^{N-1} B_2(k/N) = -(N-1)/6N$, where $B_2(T)$ is the second Bernoulli polynomial. We may assume that G is the normalizer of the diagonal subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. Also, replacing $\mathbf{a}$ by $\mathbf{a}\sigma$ with a suitable $\sigma \in \mathrm{GL}_2(\mathbb{F}_p)$, we may assume that $c = c_\infty$. Since the covering $X(p) \to X_G$ is unramified at the cusps, we have, according to item (c) of Proposition 4.3, $\mathrm{ord}_{c_\infty} w_{\mathbf{a}} = \sum_{\sigma\in G} \mathrm{ord}_{c_\infty} u_{\mathbf{a}\sigma} = 12p^2\sum_{\sigma\in G}\ell_{\mathbf{a}\sigma}$. Now we have two cases. If the entries of $\mathbf{a} = (a_1, a_2)$ are non-zero, then every non-zero element of $\mathbb{F}_p$ occurs exactly $2(p-1)$ times as the first coordinate of $\mathbf{a}\sigma$, when σ runs over G (and 0 does not occur at all). Hence, by the definition of $\ell_{\mathbf{a}}$, we have

$$\mathrm{ord}_{c_\infty} w_{\mathbf{a}} = 12p^2 \cdot 2(p-1)\sum_{k=1}^{p-1}\frac{1}{2}B_2\left(\frac{k}{p}\right) = -2p(p-1)^2 = -|G|p,$$

because $|G| = 2(p-1)^2$. And if either a_1 or a_2 is 0, then each non-zero element occurs exactly $p-1$ times, while 0 occurs $(p-1)^2$ times. Hence

$$\mathrm{ord}_{c_\infty} w_{\mathbf{a}} = 12p^2\left((p-1)\sum_{k=1}^{p-1}\frac{1}{2}B_2\left(\frac{k}{p}\right) + (p-1)^2\cdot\frac{1}{2}B_2(0)\right) = p(p-1)^3 = \frac{1}{2}|G|p(p-1).$$

Since $p \ge 3$, we have (33) in any case. □

Proof of Theorem 7.1 We argue as in the proof of Theorem 1.1. In particular, we again have w of the form $w_{\mathbf{a}}^{\pm 1}$. Further, $\Xi_1 = \log|w(P)|$ and $\Xi_2 = 0$. But now, instead of (27) we use (26), which gives

$$\log|w(P)| \le -\frac{\mathrm{ord}_c w}{p}\log|j(P)| + \log|\beta(c)| + 2|G|p \le -\frac{\mathrm{ord}_c w}{p}\log|j(P)| + 14|G|p\log p$$

(we estimate $\bigl|\log|\beta(c)|\bigr|$ by $12|G|p\log p$). Hence $\Xi_1 \le |G|\left(-\log|j(P)| + 14p\log p\right)$ due to Lemma 7.2. We again use (32), which gives $\Xi_3 \le 9|G|p$. Since $\Xi_1 + \Xi_3 = 0$, the result follows. □

8 Split Cartan Structures in the Torsion of Elliptic Curves

In this section we prove Theorems 1.3 and 1.4. We need several auxiliary results. First of all, recall the results of Masser, Wüstholz and Pellarin on the isogenies of elliptic curves. Masser and Wüstholz obtained an explicit upper bound for the degree of the minimal isogeny between two isogenous elliptic curves [11] and abelian varieties [12]; see also [13]. Pellarin [17] obtained a totally explicit version of this result for the case of elliptic curves. We use the result of Pellarin in the following form, which is a direct combination of Théorème 2 from [17] and inequality (51) on page 240 of [17].

Proposition 8.1 (Masser-Wüstholz, Pellarin). *Let E be an elliptic curve defined over a number field K of degree d. Let E' be another elliptic curve, defined over K and isogenous to E. Then there exists an isogeny $\psi : E \to E'$ of degree*

$$\deg \psi \le 10^{82} d^4 \max\{1, \log d\}^2 \left(1 + \mathrm{h}(j_E)\right)^2. \tag{34}$$

Corollary 8.2 *Let E be a non-CM elliptic curve defined over a number field K of degree d, and admitting a cyclic isogeny over K. Then the degree of this isogeny is bounded by the right-hand side of (34).*

Proof Let ϕ be a cyclic isogeny from E to E', and let $\phi^D \colon E' \to E$ be the dual isogeny. Let $\psi \colon E \to E'$ be a isogeny of degree bounded by the right-hand side of (34) which, without loss of generality, may be assumed to be cyclic. As E has no CM, the composed map $\phi^D \circ \psi$ must be multiplication by some integer, so that $\phi = \pm\psi$. □

Next we establish several simple properties of twists of elliptic curves.

Lemma 8.3 *Let E be an elliptic curve over $\mathbb{Q}$ with j-invariant $j_E \neq 0, 1728$. Then there is a twist E' of E over $\mathbb{Q}$, such that if $\ell \ge 5$ is a prime number dividing the conductor of E' then* $\mathrm{ord}_\ell\left(j_E(j_E - 1728)\right) \neq 0$.

Proof Consider the Weierstrass equation

$$y^2 + xy = x^3 - \frac{36}{j_E - 1728}x - \frac{1}{j_E - 1728}. \tag{35}$$

It is known to have discriminant $j_E^2/(j_E - 1728)^3$, and that the elliptic curve E' it defines over $\mathbb{Q}$ has j-invariant equal to j_E. It follows that E' is a twist of E over $\mathbb{Q}$ and as $j \neq 0, 1728$, this twist is necessarily quadratic. For a prime $\ell \ge 5$ such that $\mathrm{ord}_\ell\left(j_E(j_E - 1728)\right) = 0$, equation (35) defines a smooth model for E' over $\mathbb{Z}_\ell$. Therefore the minimal Weierstrass equation for E' over $\mathbb{Z}$ defines a scheme which is smooth over $\mathbb{Z}_\ell$, which means that ℓ does not divide the conductor of E'. □

We now fix a prime power p^n. Let G be a subgroup of $\mathrm{GL}(E[p^n]) \simeq \mathrm{GL}_2(\mathbb{Z}/p^n\mathbb{Z})$. We say that an elliptic curve E defined over $\mathbb{Q}$ is *endowed with a G-level structure* if the image of the natural Galois representation $\rho_{E,p^n} \colon \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(E[p^n])$ is conjugate to G.

Lemma 8.4 *Assume that G contains ± 1. Let E be an elliptic curve over $\mathbb{Q}$ with $j_E \neq 0, 1728$, endowed with a G-level structure. Then any twist of E over $\mathbb{Q}$ is endowed with a G-level structure as well.*

Proof If E' is the twist $E \otimes \chi$ of E by a character χ, and ρ_E is the Galois representation associated to the p-adic Tate module of E, then the similar object $\rho_{E'}$ for E' is the tensor product $\rho_E \otimes \chi$. Since χ has values in $\{\pm 1\} \subset G$, the curve E is endowed with a G-level structure if and only if the same is true for E'. □

The following proposition is instrumental in the proof of Theorem 1.3.

Proposition 8.5 *There exists an absolute effective constants κ such that the following holds. Let E be a non-CM elliptic curve over $\mathbb{Q}$, endowed with a structure of normalizer of split Cartan subgroup in level p^n. Then*

$$p^n \le \kappa\,(1+\mathrm{h}(j_E))^2\,. \tag{36}$$

Assuming GRH, we also have

$$p^n \le \kappa \log(N_E)(\log\log(2N_E))^6, \tag{37}$$

where N_E is the conductor of E.

Proof By the assumption, $\rho_{E,p^n}\left(\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})\right)$ is contained in the normalizer G of a split torus $G_0 \le \mathrm{GL}\left(E[p^n]\right)$. Let χ be the quadratic character of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ defined by G/G_0, and let K be the corresponding number field (which is at most quadratic over $\mathbb{Q}$). Then $\rho_{E,p^n}\left(\mathrm{Gal}(\bar{\mathbb{Q}}/K)\right)$ is contained in the split torus G_0, which implies that E admits a cyclic isogeny of degree p^n over K (and even two distinct cyclic isogenies). This implies (36) by Corollary 8.2.

Now let us assume GRH and prove (37). We apply the argument of Halberstadt and Kraus [6], which makes use of Serre's explicit version of the Chebotarev theorem [20]. Let E' be the twist of E by χ. The conductors of E' and E are equal by [6, Théorème 1]. For any prime number ℓ not dividing pN_E, the traces a_ℓ and a'_ℓ of a Frobenius substitution Frob_ℓ acting on the p-adic Tate modules of E and E' satisfy $a_\ell = a'_\ell\,\chi(\ell)$. The curve E being endowed with a K-rational isogeny of degree p, it follows from Mazur's theorem on rational isogenies [14, Theorem 1] that, if $p > 163$, we have $K \ne \mathbb{Q}$ and, consequently, $\chi \ne 1$. (One can replace here 163 by 37, because, as Mazur indicates in the introduction of [14], all curves with rational isogenies of order exceeding 37 have complex multiplication.) Since E has no complex multiplication, E and E' are not $\mathbb{Q}$-isogenous, so we have $a_\ell \ne a'_\ell$ for infinitely many ℓ. Théorème 21 from [20] implies that, assuming GRH, one finds such ℓ satisfying $\ell \le c(\log N_E)^2(\log\log 2N_E)^{12}$. Since $a_\ell \ne a'_\ell$, we have $a_\ell \ne 0$ and $\chi(\ell) = -1$, which means that $\rho_{E,p^n}\left(\mathrm{Frob}_\ell\right)$ belongs to $G \setminus G_0$. Since all elements from $G \setminus G_0$ have trace 0, we obtain $p^n \mid a_\ell$. Now Hasse's bounds imply that $p^n \le |a_\ell| \le 2\sqrt{\ell}$, which yields (37). $\square$

Proof of Theorem 1.3 Assume that $X_{\mathrm{split}}(p^n)(\mathbb{Q})$ has a non-CM and non-cuspidal point P. We want to show that, for sufficiently large p, we have $n \le 2$, and even $n \le 1$ assuming GRH.

Our point P gives rise to a non-CM elliptic curve E. Since P induces a point in $X_{\mathrm{split}}(p)(\mathbb{Q})$, the results of Momose and Merel [15, Theorem 3.1] imply that either $p \le 13$ or $j(P) = j_E$ belongs to $\mathbb{Z}$. Now Theorem 7.1 yields

$$\log|j_E| \le 23p\log p, \tag{38}$$

which, together with (36) gives $p^n \le c(p\log p)^2$ for some constant c (since $j_E \in \mathbb{Z}$ we have $\mathrm{h}(j_E) = \log|j_E|$). Hence $n \le 2$ for sufficiently large p, proving the first (unconditional) statement.

Now let us prove the second statement. Lemmas 8.3 and 8.4 allow us to assume (replacing E, if necessary, by a quadratic twist) that every prime $\ell \ge 5$ dividing N_E, divides either j_E or $j_E - 1728$. The curve E has potential good reduction at all primes, so $\mathrm{ord}_\ell(N_E) = 2$, and the exponents of the conductor at 2 and 3 are at most 8 and 5 respectively. Therefore $N_E \le 2^8 \cdot 3^5 \cdot j_E^2(j_E - 1728)^2$. Combining this with (37) and (38), we obtain, assuming GRH, that $p^n \le cp(\log p)^7$ for some constant c. Therefore $n \le 1$ for sufficiently large p. $\square$

Proof of Theorem 1.4 The proof is very similar to that of the first part of Theorem 1.3. Let p, q and r be distinct split Cartan deficient primes for a non-CM elliptic curve $E/\mathbb{Q}$. We assume $11 \le p < q < r$. Again applying the results of Momose and Merel, we obtain $j_E \in \mathbb{Z}$. Hence E gives rise to a point on $Y_{\mathrm{split}}(p)(\mathbb{Z})$, and Theorem 7.1 yields (38).

On the other hand, over some quadratic field E admits a cyclic isogeny of degree p, and the same is true for q and r. Hence over some field of degree (at most) 8 the curve E admits a cyclic isogeny of degree pqr. Using Corollary 8.2 and (38) we obtain $p^3 \le pqr \le c(p\log p)^2$, which is impossible when p exceeds certain p_0. $\square$

References

[1] *Yu. Bilu, P. Parent,* Integral j-invariants and Cartan structures for elliptic curves, C. R. Acad. Sci. Paris, Ser. I **346** (2008), 599–602.

[2] *E. Bombieri*, On Weil's "théorème de décomposition", Amer. J. Math. **105** (1983), 295–308.

[3] *A. C. Cojocaru*, On the surjectivity of the Galois representations associated to non-CM elliptic curves (with an appendix by Ernst Kani), Canad. Math. Bull. **48** (2005), 16–31.

[4] *A. C. Cojocaru, C. Hall*, Uniform results for Serre's theorem for elliptic curves, Int. Math. Res. Not. 2005, 3065–3080.

[5] *P. Deligne, M. Rapoport*, Les schémas de modules de courbes elliptiques, in "Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972)", pp. 143–316; Lecture Notes in Math. **349**, Springer, Berlin, 1973.

[6] *E. Halberstadt, A. Kraus*, Sur les modules de torsion des courbes elliptiques, Math. Annalen **310** (1998), 47–54.

[7] *D. S. Kubert, S. Lang*, Modular units, Grund. math. Wiss. **244**, Springer, New York-Berlin, 1981.

[8] *S. Lang*, Elliptic Functions, Addison-Wesley, 1973.

[9] *S. Lang*, Introduction to modular forms, Grund. math. Wiss. **222**, Springer, Berlin-New York, 1976.

[10] *A. Levin*, Variations on a theme of Runge: effective determination of integral points on certain varieties, submitted.

[11] *D. W. Masser, G. Wüstholz*, Estimating isogenies on elliptic curves, Invent. Math. **100** (1990), 1–24.

[12] *D. W. Masser, G. Wüstholz*, Isogeny estimates for abelian varieties and finiteness theorems, Ann. of Math. (2) **137** (1993), 459–472.

[13] *D. W. Masser, G. Wüstholz*, Galois properties of division fields of elliptic curves, Bull. London Math. Soc. **25** (1993), 247–254.

[14] *B. Mazur*, Rational isogenies of prime degree (with an appendix by D. Goldfeld). Invent. Math. **44** (1978), 129–162.

[15] *L. Merel*, Normalizers of split Cartan subgroups and supersingular elliptic curves, in "Diophantine Geometry" (edited by U. Zannier), pp. 237–255; CRM Series **4**, Edizioni della Normale, Pisa, 2007.

[16] *P. Parent*, Towards the triviality of $X_0^+(p^r)(\mathbb{Q})$ for $r > 1$, Compos. Math. **141** (2005), 561–572.

[17] *F. Pellarin*, Sur une majoration explicite pour un degré d'isogénie liant deux courbes elliptiques, Acta Arith. **100** (2001), 203–243.

[18] *M. Rebolledo*, Module supersingulier, formule de Gross-Kudla et points rationnels de courbes modulaires, Pacific J. Math. **234** (2008), 167–184.

[19] *J.-P. Serre*, Propriétés galoisiennes des points d'ordre fini des courbes elliptiques, Invent. Math. **15** (1972), 259–331.

[20] *J.-P. Serre*, Quelques applications du théorème de densité de Chebotarev, Publ. Math. IHES **54** (1981), 323–401.

[21] *G. Shimura*, Introduction to the arithmetic theory of automorphic functions, Publ. Math. Soc. Japan **11**, Iwanami Shoten, Tokyo; Princeton University Press, Princeton, N.J., 1971.